SVDs in 3D and Beyond

Vince Fernando
Division of Structural Biology
Wellcome Trust Centre for Human Genetics
University of Oxford

20 July 2004

vince at strubi dot ox dot ac dot uk

Norms

$$
\begin{aligned}
\|\mathbf{y}\|_2 &= \sqrt{\{|y_1|^2 + |y_2|^2 + \ldots + |y_n|^2\}} \\
\|A\|_2 &= \max_{\|\mathbf{y}\|=1} \|A\mathbf{y}\|_2 \quad \text{2-norm} \\
\|A\|_{\mathsf{F}}^2 &= \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \quad \text{Frobenius norm} \\
\|A\| &= \text{either } \|A\|_2 \text{ or } \|A\|_{\mathsf{F}}
\end{aligned}
$$

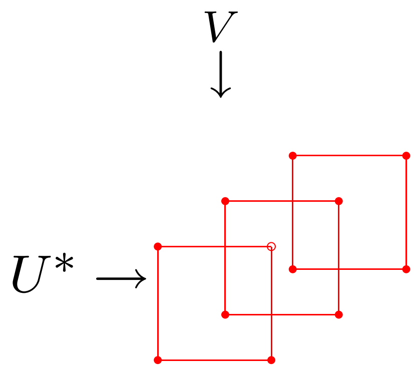
Classical Tensor Notation

- too many indices
- too many summations
- multiplications can be tedious
- can be clumsy
- far away from matrix notation

A 3D Array

- a 3D array can be sliced
- each slice is a matrix (2D array)
- each matrix = a set of data
- each matrix = a cross-section of data
- each matrix = an image

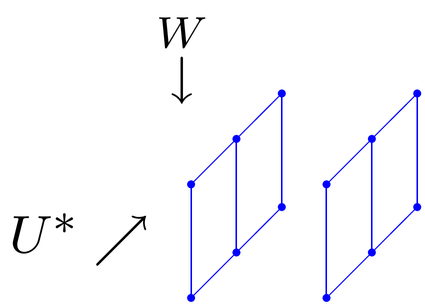
Slicing an $m \times n \times p$ Array Θ



$m = n = 2 \ , \ p = 3$

- $\Theta(:,:,k)$ slices $k = 1 : p$
- U^* is applied to all (np) columns
- V is applied to all (mp) rows
- U^*, V multiplications are well defined
- multiplications by matrix inverses are well defined

Slicing an $m \times n \times p$ Array Θ



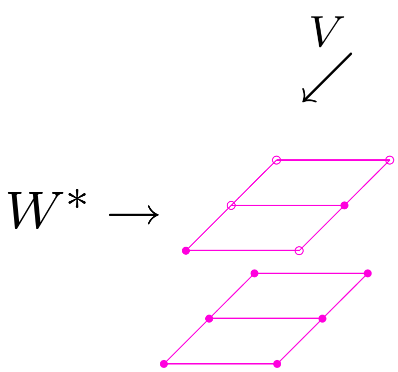
$$m = n = 2 \ , \ \ p = 3$$

- $\Theta(:, j, :)$ slices $j = 1 : n$

- U^* is applied to all (pn) columns

- W is applied to all (mn) rows

Slicing an $m \times n \times p$ Array $\ominus$



$$m = n = 2 \;\;,\;\; p = 3$$

- $(i, :, :)$ slices $i = 1 : m$

- W^* is applied to all (mn) columns

- V is applied to all (mp) rows

Tensor Product $\diamond$

A be a matrix
w be a vector

Define a tensor (3D array)

$$\Gamma = \mathbf{w} \diamond A$$

qualified by

$$\gamma_{ijk} \quad = \quad a_{ij} w_k$$

Alternatively,

$$\gamma_{ijk} \quad = \quad b_{ik} v_j \quad , \quad \Gamma = \mathbf{v} \diamond B$$

$$\gamma_{ijk} \quad = \quad c_{jk} u_i \quad , \quad \Gamma = \mathbf{u} \diamond C$$

Vectorizing Operator ν

Kronecker product: $\otimes$

$$
\begin{aligned}
\Gamma &= \mathbf{w} \diamond A \\
\nu_3\{\Gamma\} &= \mathbf{w} \otimes \nu_2\{A\} \\
\Gamma &= \nu_3^{-1}\{\mathbf{w} \otimes \nu_2\{A\}\}
\end{aligned}
$$

- **One tensor product $\rightarrow$ class 1**
- **$\nu\{.\}$ conserves all elements of $\{.\}$**
- **no information loss**
- **rearranged outer product**
- **inverse of ν is well defined**
- **an operator denoted by a single letter**

Flattening Operator $\flat$

$$
\begin{aligned}
\Gamma &= \mathbf{w} \diamond A \\
\flat\{\Gamma\} &= \mathbf{w} \otimes \nu\{A\}^t \\
\Gamma &= \flat^{-1}\{\mathbf{w} \otimes \nu\{A\}^t\}
\end{aligned}
$$

- $\flat$ conserves all elements of Γ
- no information loss
- inverse of $\flat$ is well defined

Tensor versus Kronecker

- $\otimes$ creates vectors out of vectors
- $\otimes$ creates matrices out of vectors
- $\otimes$ creates matrices out of matrices
- $\diamond$ creates new dimensions

An Example

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad , \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

$$\begin{aligned} \Gamma &= \mathbf{w} \diamond A \\ \gamma_{ijk} &= a_{ij} w_k \end{aligned}$$

Cross-sections wrt to the third dimension

$$\begin{aligned} \Gamma(:,:,1) &= w_1 A \\ \Gamma(:,:,2) &= w_2 A \\ \Gamma(:,:,3) &= w_3 A \end{aligned}$$

These are the natural cross-sections

Example Continued

$$\flat(\Gamma) = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & a_{12} & a_{22} \end{bmatrix}$$

- $\nu(\Gamma)$ is a tall vector
- $\flat(\Gamma)$ is a rank-1 matrix

Cross-sections wrt Second Dimension

$$\begin{aligned}\Gamma(:,1,:) &= \begin{bmatrix} w_1 a_{11} & w_2 a_{11} & w_3 a_{11} \\ w_1 a_{21} & w_2 a_{21} & w_3 a_{21} \end{bmatrix}, \\ &= \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \mathbf{w}^t\end{aligned}$$

$$\begin{aligned}\Gamma(:,2,:) &= \begin{bmatrix} w_1 a_{12} & w_2 a_{12} & w_3 a_{12} \\ w_1 a_{22} & w_2 a_{22} & w_3 a_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \mathbf{w}^t\end{aligned}$$

- γ_{111} on the top left of $\Gamma(:,1,:)$
- reduced ranks in these cross-sections

Cross-sections wrt First Dimension

$$\Gamma(1,:,:) = \begin{bmatrix} w_1 a_{11} & w_2 a_{11} & w_3 a_{11} \\ w_1 a_{12} & w_2 a_{12} & w_3 a_{12} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} \mathbf{w}^t$$

$$\Gamma(2,:,:) = \begin{bmatrix} w_1 a_{21} & w_2 a_{21} & w_3 a_{21} \\ w_1 a_{22} & w_2 a_{22} & w_3 a_{22} \end{bmatrix} = \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} \mathbf{w}^t$$

- γ_{111} on the top left of $\Gamma(1,:,:)$
- reduced ranks in these cross-sections

Cross-sections

- **implementation dependent**
- **we follow MATLAB and Fortran**
- **trivially different from De Lathauwer et al (2000)**
- $(1, 1, 1)$ **element at the top left**

Flattening with a Template

Arbitrary $\Theta \in \mathcal{C}^{m\times n\times p}$
Template $\Gamma \in \mathcal{C}^{m\times n\times p}$

$$\begin{aligned} \Gamma &= \mathbf{w} \diamond A \\ \flat\{\Gamma\} &= \mathbf{w} \otimes A^t \end{aligned}$$

Using the same element mapping rules

$$\begin{aligned} \flat\{\Theta\} &= F \\ \flat^{-1}\{F\} &= \Theta \end{aligned}$$

An Example

$$\flat\{\Gamma\} = \mathbf{w}\left[\begin{array}{cccc} a_{11} & a_{21} & a_{12} & a_{22} \end{array}\right]$$

$$\flat\{\Theta\} = \left[\begin{array}{cccc} \theta_{111} & \theta_{211} & \theta_{121} & \theta_{221} \\ \theta_{112} & \theta_{212} & \theta_{122} & \theta_{222} \\ \theta_{113} & \theta_{213} & \theta_{123} & \theta_{223} \end{array}\right]$$

Another 3D Class

a, b and c be vectors

We can define

$$\Gamma = \mathbf{a} \diamond \mathbf{b} \diamond \mathbf{c}$$

where

$$\gamma_{ijk} = a_i b_j c_k.$$

Two tensor products $\rightarrow$ class 2

Γ is rank 1

The Two 3D Classes

- Class 1 $\rightarrow$ Tucker/multilinear type model

- Class 2 $\rightarrow$ parafac/candecomp type model

Tucker (1966)
multilinear: De Lathauwer et al (2000)

parafac: Harshman (1970)
candecomp: Carroll and Chang (1970)

Orthogonal Vectors

$$
\begin{aligned}
U &= [\mathbf{u}^{(1)} \ldots \mathbf{u}^{(m)}] \quad , \quad U^*U = I \\
V &= [\mathbf{v}^{(1)} \ldots \mathbf{v}^{(n)}] \quad , \quad V^*V = I \\
W &= [\mathbf{w}^{(1)} \ldots \mathbf{w}^{(p)}] \quad , \quad W^*W = I
\end{aligned}
$$

In general, U, V and W are unitary

Three Linear Models for the Same Array

$$\Gamma \in \mathcal{C}^{m \times n \times p}$$

$$\Gamma \;=\; \sum_{k=1}^{p} \mathbf{w}^{(k)} \diamond A^{(k)}$$

$$\Gamma \;=\; \sum_{j=1}^{n} \mathbf{v}^{(j)} \diamond B^{(j)}$$

$$\Gamma \;=\; \sum_{i=1}^{m} \mathbf{u}^{(i)} \diamond C^{(i)}$$

- tensor summations

- Reminds one-sided Jacobi algorithms

Tensor from Core

The core array $\Sigma \in \mathcal{C}^{m \times n \times p}$

$$\Gamma_1 = \sum_{k=1}^{p} \mathbf{w}^{(k)} \diamond \Sigma(:,:,k)$$

$$\Gamma_2 = \sum_{j=1}^{n} \mathbf{v}^{(j)} \diamond \Gamma_1(:,j,:)$$

$$\Gamma = \sum_{i=1}^{m} \mathbf{u}^{(i)} \diamond \Gamma_2(i,:,:)$$

- tensor multiplications in any order
- can be extended to 4D and higher

Core from Tensor

Let $\widehat{\mathbf{w}}$ be the columns of $W^{-1} = W^*$

$$\Gamma_1 = \sum_{k=1}^{p} \widehat{\mathbf{w}}^{(k)} \diamond \Gamma(:,:,k)$$

$$\Gamma_2 = \sum_{j=1}^{n} \widehat{\mathbf{v}}^{(j)} \diamond \Gamma_1(:,j,:)$$

$$\Sigma = \sum_{i=1}^{m} \widehat{\mathbf{u}}^{(i)} \diamond \Gamma_2(i,:,:)$$

- tensor multiplications in any order

Core Orthogonality

$$\text{sum}\{\bar{\Sigma}(i_1, :, :) \circ \Sigma(i_2, :, :)\} = \alpha_1 \delta_{i_1 i_2}$$

$$\text{sum}\{\bar{\Sigma}(:, j_1, :) \circ \Sigma(:, j_2, :)\} = \alpha_2 \delta_{j_1 j_2}$$

$$\text{sum}\{\bar{\Sigma}(:, :, k_1) \circ \Sigma(:, :, k_2)\} = \alpha_3 \delta_{k_1 k_2}$$

Hadamard (Schur) product $\circ$
Element by element product $\circ$
Kronecker delta δ_{ij}

Core Orthogonality without $\circ$

$$\nu\{\Sigma(i_1,:,:)\}^*\nu\{\Sigma(i_2,:,:)\} \;=\; \alpha_1\delta_{i_1i_2}$$

$$\nu\{\Sigma(:,j_1,:)\}^*\nu\{\Sigma(:,j_2,:)\} \;=\; \alpha_2\delta_{j_1j_2}$$

$$\nu\{\Sigma(:,:,k_1)\}^*\nu\{\Sigma(:,:,k_2)\} \;=\; \alpha_3\delta_{k_1k_2}$$

Core Orthogonality

- Core in matrix SVD is diagonal

- Tucker model core is NOT diagonal

- Columns (rows) of matrix core are orthogonal

- Tucker model core slices are orthogonal

Tucker Model: Questions

Matrix SVD is optimal
Is Tucker model optimal ?

Explain core orthogonality

Best Rank One Approximation

Eckart-Young-Mirsky Problem

$$\min_{x,y} \|F - xy^*\|$$

$$\begin{aligned} \mathbf{x} &= \alpha\phi \\ \mathbf{y} &= (\sigma/\alpha)\psi \end{aligned}$$

- ϕ first left singular vector of F
- ψ first right singular vector of F
- σ first singular value of F
- $\alpha (= 1)$ arbitrary positive value
- similarly, the second best approximation

Best Rank k Approximation

Eckart-Young-Mirsky Problem

$$\min_{X,Y} \|F - XY^*\|$$

$$\begin{aligned} \mathbf{x}^{(i)} &= \phi^{(i)} \\ \mathbf{y}^{(i)} &= \sigma_i \psi^{(i)} \\ X &= [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}] \\ Y &= [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}] \end{aligned}$$

Tucker Model Algorithm

Arbitrary $\Theta \in \mathcal{C}^{m \times n \times p}$
Template $\Gamma = \mathbf{w} \diamond A$

$F = \flat\{\Theta\}$
Find $\mathbf{w}, A$ by $\quad \min \|F - \mathbf{w}\nu(A)^t\|$
number of solutions $= p$

$\Theta = \sum_{k=1}^{p} \mathbf{w}^{(k)} \diamond A^{(k)}$

Core Orthogonality

Due to the orthogonality of

$$A^{(k)}; \quad , \quad k = 1 : p$$

Tucker Singular values

- singular values of F for each template
- three possible templates in 3D

Complete Orthogonal Parafac

This is a decomposition

Not a model

Could be the SVD in 3D

Largest Singular Value of a Matrix

$$\sigma = \|A\|_2 \quad \text{primary definition}$$

$$= \max_{\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1} \mathbf{x}^t A \mathbf{y}$$

$$\sigma^2 = \max_{\|\mathbf{y}\|_2 = 1} \mathbf{y}^t A^* A \mathbf{y}$$

$$= \max_{\|\mathbf{x}\|_2 = 1} \mathbf{x}^t A A^* \mathbf{x}$$

$\mathbf{x} \in \mathcal{C}^m \;,\; \mathbf{y} \in \mathcal{C}^n$

Largest Singular Value of a 3D Array ?

$\mathbf{x} \in \mathcal{C}^m \ , \ \mathbf{y} \in \mathcal{C}^n \ , \ \mathbf{z} \in \mathcal{C}^p$

$$\begin{aligned}
\sigma_{3D} &= \max_{\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = \|\mathbf{z}\|_2 = 1} \sum_{k=1}^{p} z_k \mathbf{x}^t \Theta(:,:,k) \mathbf{y} \\
&= \max_{\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = \|\mathbf{z}\|_2 = 1} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \theta_{ijk} x_i y_j z_k
\end{aligned}$$

This is a reasonable secondary definition

see Zhang and Golub (2001)

What is the primary definition ?

Sets of Matrices

$$\begin{aligned}
\mathcal{A} &= \{A_1, A_2, \ldots A_p\} \\
\mathcal{B} &= \{B_1, B_2, \ldots B_p\} \\
\alpha\mathcal{A} &= \{\alpha A_1, \alpha A_2, \ldots \alpha A_p\} \\
\mathcal{A} + \mathcal{B} &= \{A_1 + B_1, A_2 + B_2, \ldots A_p + B_p\}
\end{aligned}$$

A Norm for a Set of Matrices

Define the function μ,

$$\mu(\mathcal{A}) \;=\; \max_{\|\mathbf{z}\|_2=1} \|\sum_{k=1}^{p} z_k A_k\|_2$$

Then

$$\mu(\mathcal{A}) \;\geq\; 0$$

$$\mu(\mathcal{A}) \;=\; 0 \;\rightarrow \mathcal{A} = 0$$

$$\mu(\alpha\mathcal{A}) \;=\; |\alpha|\mu(\mathcal{A})$$

$$\mu(\mathcal{A}+\mathcal{B}) \;\leq\; \mu(\mathcal{A}) + \mu(\mathcal{B})$$

Thus, $\mu(.)$ is a norm induced by the 2-norm

A Norm for a 3D Array

Can be defined via the slices

$$\begin{aligned}
\mu(\Theta) &= \max_{\|\mathbf{z}\|_2=1} \| \sum_{k=1}^{p} z_k \Theta(:,:,k) \|_2 \\
&= \max_{\|\mathbf{y}\|_2=1} \| \sum_{j=1}^{m} y_k \Theta(:,j,:) \|_2 \\
&= \max_{\|\mathbf{x}\|_2=1} \| \sum_{i=1}^{n} x_k \Theta(i,:,:) \|_2 \\
&= \max_{\|\mathbf{x}\|_2=\|\mathbf{y}\|_2=\|\mathbf{z}\|_2=1} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \theta_{ijk} x_i y_j z_k
\end{aligned}$$

A Bound for the Norm

$$\mu(\Theta) \leq \sqrt{\sum_{k=1}^{p} \|\Theta(:,:,k)\|_2^2}$$

$$\mu(\Theta) \leq \sqrt{\sum_{j=1}^{n} \|\Theta(:,j,:)\|_2^2}$$

$$\mu(\Theta) \leq \sqrt{\sum_{i=1}^{m} \|\Theta(i,:,:)\|_2^2}$$

Invariant Property

$$
\begin{aligned}
\mu\left(\sum_{k=1}^{p} \mathbf{w}^{(k)} \diamond \Theta(:,:,k)\right) &= \mu(\Theta) \\
\mu\left(\sum_{j=1}^{n} \mathbf{v}^{(j)} \diamond \Theta(:,j,:)\right) &= \mu(\Theta) \\
\mu\left(\sum_{i=1}^{m} \mathbf{u}^{(i)} \diamond \Theta(i,:,:)\right) &= \mu(\Theta)
\end{aligned}
$$

where U, V and W are unitary

Three Subproblems

$$\mu(\Theta) \;=\; \max_{\|\mathbf{z}\|_2=1} \| \sum_{k=1}^{p} z_k \Theta(:,:,k) \|_2$$

- μ_{cc} $\leftarrow$ **Θ and $\mathbf{z}$ are complex (generic)**

- μ_{rc} $\leftarrow$ **Θ is real and $\mathbf{z}$ is complex**

- μ_{rr} $\leftarrow$ **Θ and $\mathbf{z}$ are real**

A Surprising Result

$$\mu_{rr}(\Theta) \;\leq\; \mu_{rc}(\Theta)$$

The following statements are wrong

- Without loss of generality we assume that the array is real

- This can be easily extended to the complex case

An Algorithm for the Largest Singular Value

$\Theta \in \mathcal{C}^{m \times m \times m}$

Use permutations to make $|\theta_{111}|$ the largest (permutation matrices are unitary)

$[U, S, V] = svd(\Theta(:, :, 1))$
Apply U^* and V to Θ

$[U, S, W] = svd(\Theta(:, 1, :))$
Apply U^* and W to Θ

$[W, S, V] = svd(\Theta(1, :, :))$
Apply W^* and V to Θ

Repeat above until convergence

Convergence

Element θ_{111} does not increase

The three edges

$\Theta(1, 1, :)$
$\Theta(1, :, 1)$
$\Theta(:, 1, 1)$

are zero except for the element θ_{111}

An Algorithm for the Second Singular Value

Use permutations to make $|\theta_{222}|$ the largest without disturbing θ_{111}

$[U, S, V] = svd(\Theta(2:m, 2:m, 2))$
Apply U^* and V to Θ

$[U, S, W] = svd(\Theta(2:m, 2, 2:m))$
Apply U^* and W to Θ

$[W, S, V] = svd(\Theta(2, 2:m, 2:m))$
Apply W^* and V to Θ

Repeat above until convergence

Convergence

Element θ_{222} does not increase

The three edges

$\Theta(2, 2, 3 : m)$
$\Theta(2, 3 : m, 2)$
$\Theta(3 : m, 2, 2)$

are zero

The 3D SVD

For $\Theta \in \mathcal{C}^{m\times m\times m}$

$$\Theta = \sum_{i}^{m}\sum_{j}^{m}\sum_{k}^{m} \pi_{i,j,k}\mathbf{u}^{(i)} \diamond \mathbf{v}^{(j)} \diamond \mathbf{w}^{(k)}$$

where

$\pi_{q,q,q}$
for $q = 1 : m$ are the 'singular values'

The edges:

$\pi_{q+1:m,q,q} = 0$
$\pi_{q,q+1:m,q} = 0$
$\pi_{q,q,q+1:m} = 0$

for $q = 1 : m-1$

Main Properties

$$\begin{aligned}
\pi_{1,1,1} &\geq \pi(1, j, k) \\
&\geq \pi(i, 1, k) \\
&\geq \pi(i, j, 1)
\end{aligned}$$

Main Properties

$$\begin{aligned}\pi_{1,1,1} &\geq \sigma\{\Pi(1, 2:m, 2:m)\} \\ &\geq \sigma\{\Pi(2:m, 1, 2:m)\} \\ &\geq \sigma\{\Pi(2:m, 2:m, 1)\}\end{aligned}$$

$\pi_{2,2,2}$

Peel off the first layer

Use properties of $\pi_{1,1,1}$

Monotonicity

$$\pi_{1,1,1} \geq \pi_{2,2,2} \geq \pi_{3,3,3} \geq \ldots$$

Orthogonal Rank ?

Number of non-zero $\pi_{q,q,q}$

Conclusions: Tucker Model

- 3D Tucker is based on 3 SVDs
- Tucker model is optimal
- Three ways to express the same model
- Eckart-Young-Mirsky (EYM) in 3 ways
- Core orthogonality due to EYM
- only one tensor (outer) product
- one-sided Jacobi expansions

Conclusions: 3D Norm

- A norm induced by the vector 2-norm
- Compatible with Zhang-Golub
- Computable bounds for the norm
- Generic problem is complex

Conclusions: 3D SVD

- Based recursively on the 3D norm
- Complete orthogonal parafac
- Two tensor products
- Edges are zero
- Diagonals give 'ordered singular values'
- Off-diagonals give 'sub-singular values'

Conclusions: Tensor Classes

- Tucker is class 1

- 3D parafac is class 2

- Vector spaces are vectors, matrices 3D arrays etc

- Tensors are products of vector spaces

Acknowledgements

- Professor Gene Golub
- Professor Stephen Fuller
- Dr Karin Anduleit

Acknowledgements

- Wellcome Trust
- University of Oxford
- American Institute of Mathematics
- National Science Foundation